



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/563,311	12/30/2005	Eiji Murakami	96790P517	6897
8791 7590 02/05/2008 BLAKELY SOKOLOFF TAYLOR & ZAFMAN 1279 OAKMEAD PARKWAY SUNNYVALE, CA 94085-4040			EXAMINER HWA, SHYUE JIUNN	
			ART UNIT 2163	PAPER NUMBER
			MAIL DATE 02/05/2008	DELIVERY MODE PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary

Application No.

10/563,311

Applicant(s)

MURAKAMI ET AL.

Examiner

James Hwa

Art Unit

2163

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 21 November 2007.
- 2a) ☒ This action is **FINAL**. 2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1,3-6,8-10,12-15,17 and 18 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1, 3-6, 8-10, 12-15, 17 and 18 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
- ☐ Certified copies of the priority documents have been received.
 - ☐ Certified copies of the priority documents have been received in Application No. _____.
 - ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- ☒ Notice of References Cited (PTO-892)
- ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- ☐ Information Disclosure Statement(s) (PTO/SB/08)
Paper No(s)/Mail Date _____
- ☐ Interview Summary (PTO-413)
Paper No(s)/Mail Date _____
- ☐ Notice of Informal Patent Application
- ☐ Other: _____

DETAILED ACTION

1. Applicant has amended claims 1, 3-6, 8-10, 12-15, 17 and 18 in the amendment filed on 11/21/2007. Claims 2, 7, 11 and 16 are canceled. Claims 1, 3-6, 8-10, 12-15, 17 and 18 are pending in this Office Action.

Response to Arguments

2. Applicant's arguments in regards to the rejections to claims 1 and 10 under 35 U.S.C. 103(a), have been fully considered. Applicant seems to argue the claims as amended. Consequently, new grounds of rejection are set forth below as necessitated by Applicant's amendment.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was

not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g) prior art under 35 U.S.C. 103(a).

3. Claims 1, 3-6, 8-10, 12-15 and 17-18 are rejected under 35 U.S.C. 103(a) as being unpatentable over Tokuda et al. (US Patent No. 7,024,400 B2, hereinafter "Tokuda") in view of Glover (US Patent Application No. 2003/0221163 A1, hereinafter "Glover") and Agrawal et al. (US Patent Application No. 2001/0037324 A1, hereinafter "Agrawal").

As to claims 1 and 10, Tokuda teaches the claimed limitations:

"A sentence classification device characterized" as document classification is important not only in office document processing but also in implementing an efficient information retrieval system (column 1, lines 13-15).

"A term list having a plurality of terms each comprising not less than one word" as a term is defined as a word or a phrase that appears in at least two documents (column 4, lines 5-6).

"DT matrix generation module for generating a DT matrix two-dimensionally expressing a relationship between each document contained in a document set and said each term" as the term by document matrix of the original documents (column 9, lines 23; see also table 1).

"DT matrix transformation module for generating a transformed DT matrix having clusters having blocks of associated documents by transforming the DT matrix obtained by said DT matrix generation module on the basis of a DM decomposition method" as

exploiting the singular vector decomposition method, the major left singular vectors associated with the largest singular values are selected as a major vector space called an intra-DLSI space, or an I-DLSI space (column 3, lines 2-5). The extra-DLSI space or the E-DLSI space can similarly be obtained by setting up a differential term by extra-document matrix where each column of the matrix denotes a differential document vector between the document vector and the centroid vector of the cluster, which does not include the document. The extra-DLSI space may then be constructed by the major left singular vectors associated with the largest singular values (column 3, lines 18-25).

“Classification generation module for generating classifications associated with the document set on the basis of a relationship between each cluster on the transformed DT matrix obtained by said DT matrix transformation module and said each document classified according to the clusters” as given a new document to be classified, a best candidate cluster to be recalled from the clusters can be selected from among those clusters having the highest probabilities of being the given differential intra-document vector (column 3, lines 10-13).

The differences in word usage between the document and a cluster's centroid vector, the differential document vector is capable of capturing the relation between the particular document and the cluster (Column 2, lines 41-46).

Tokuda does not explicitly teach the claimed limitation “wherein the classification generation module comprises a virtual representative document generation module for generating a virtual representative document, for each cluster on a transformed DT matrix, from a term of each document belonging to the cluster”

Glover teaches using a virtual document comprising extended anchor text to determine whether a web page is to be classified into a given category (page 2, paragraph 0013).

Generating a classification output of the target web page utilizing a trained full-text classifier; and combining the classification output of the virtual document classifier and the classification output of the full-text classifier to generate a combined classification output for the target web page (page 2, paragraph 0018).

Tokuda does not explicitly teach the claimed limitation "large classification generation module for generating a large classification of documents from each document in a bottom-up manner by repeatedly performing hierarchical clustering processing of setting a DT matrix generated by said DT matrix generation module in an initial state, causing said virtual representative document generation module to generate a virtual representative document for each cluster on a transformed DT matrix generated from the DT matrix by said DT matrix transformation module, generating a new DT matrix used for next hierarchical clustering processing by adding the virtual representative document to the transformed DT matrix and deleting documents belonging to the cluster of the virtual representative document from the transformed DT matrix, and outputting, for said each cluster, information associated with the documents constituting the cluster as large classification data".

Agrawal teaches for organizing a large text database into a hierarchy of topics and for maintaining this organization as documents are added and deleted and as the topic hierarchy changes (abstract). For such classifiers, feature sets larger than 100 are

considered extremely large. Document classification may require more than 50,000 (page 2, paragraph 0019).

Singular value decomposition on the term-document matrix has been found to cluster semantically related documents together even if they do not share keywords (page 2, paragraph 0021).

The feature set changes by context as the classification process proceeds down the taxonomy. As a result, jargon common to lower nodes of the taxonomy are filtered out and the classification accuracy remains high in spite of the reduction in the number of terms and candidate classes inspected (page 3, paragraph 0029).

Each document in the database has been pre-classified. The user may then enter a command through the user input device to cause the system to select at least one of the displayed sub-topics. This process is repeated as necessary to refine the query topic until the user's information need is satisfied (page 5, paragraph 0084).

A parent class inherits, in an additive fashion, the statistics of its children, since each training document generates rows for each topic node from the assigned topic up to the root (page 13, paragraph 0204).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made, having the teachings of Tokuda, Glover and Agrawal before him/her, to modify Tokuda a DM decomposition method used in a graph theory because that would improve the document search performance include speed and accuracy as taught by Agrawal (page 14, paragraph 0216).

As to claims 3 and 12, Tokuda teaches the claimed limitations:

“Characterized by further comprising label generation module for outputting each term strongly connected to each document belonging to said arbitrary cluster as a label indicating a classification of the cluster” as a new efficient supervised document classification procedure introduced, whereby learning from a given number of labeled documents preclassified into a finite number of appropriate clusters in the database, the classifier developed will select and classify any of new documents introduced into an appropriate cluster within the classification stage (column 2, lines 21-25).

As to claims 4 and 13, although Tokuda teaches the extra-DLSI space, or the E-DLSI space can similarly be obtained by setting up a differential term by extra-document matrix where each column of the matrix denotes a differential document vector between the document vector and the centroid vector of the cluster which does not include the document (column 3, lines 18-23).

Tokuda does not explicitly teach the claimed limitation:

“Characterized by further comprising document organization module for sequentially outputting documents belonging to said arbitrary cluster or all documents in an arrangement order of the documents in the transformed DT matrix”.

Agrawal teaches given $k^*(c)$, the sorted Fisher table is scanned while copying the first $k^*(c)$ rows for the run corresponding to class c to an output table and discarding the remaining terms. This involves completely sequential IO (page 12, paragraph 0187).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made, having the teachings of Tokuda, Glover and Agrawal before him/her, to modify Tokuda the document organization because that would improve the document search performance include speed and accuracy as taught by Agrawal (page 14, paragraph 0216).

As to claims 5 and 14, Tokuda teaches the claimed limitations:

“Characterized by further comprising summary generation module for outputting, as a summary of said arbitrary document, a sentence of sentences constituting the document which contains a term strongly connected to the document” as the setting up of a DLSI space-based classifier is summarized. Documents are preprocessed, to identify and distinguish terms, either of the word or noun phrase, from stop words. System terms are then constructed, by setting up the term list as well as the global weights. The process continues with normalization of the document vectors, of all the collected documents, as well as the centroid vectors of each cluster. Following document vector normalization, the differential term by document matrices may be constructed by intra-document or extra-document construction (column 7, lines 24-34).

As to claims 6 and 15, Tokuda teaches the claimed limitations:

“Characterized by further comprising: term list edition module for adding or deleting an arbitrary term with respect to the term list; and index generation module for making said DT matrix generation module generate DT matrices by using term lists

before and after edition by said term list edition module, and generating and outputting an index indicating validity of the edition from the DT matrices” as the Latent Semantic Indexing (LSI) with Singular Value Decomposition (SVD) has proved to be a most efficient method for the dimensionality reduction scheme in document analysis and extraction, providing a powerful tool for the classifier when introduced into document retrieval with a good performance confirmed by empirical studies. A distinct advantage of LSI-based dimensionality reduction lies in the fact that among all the projections on all the possible space having the same dimensions, the projection of the set of document vectors on the LSI space has a lowest possible least-square distance to the original document vectors. This implies that the LSI finds an optimal solution to dimensional reduction. In addition to the role of dimensionality reduction, the LSI with SVD also is effective in offering a dampening effect of synonymy and polysemy problems with which a simple scheme of deleting terms cannot be expected to cope. Also known as a word sense disambiguation problem, the source of synonymy and polysemy problems can be traced to inherent characteristics of context sensitive grammar of any natural language (column 1, line 57 to column 2, line 9).

Tokuda does not explicitly teach the claimed limitation “edition means for adding or deleting an arbitrary term with respect to the term list”.

Agrawal teaches a system, process, and article of manufacture for organizing a large text database into a hierarchy of topics and for maintaining this organization as documents are added and deleted and as the topic hierarchy changes (abstract)

Addition and deletion of documents to given topics, as well as reorganization of the topic hierarchy itself, are easily handled. The text models built at each node also yield a means to summarize a number of documents using a few descriptive keywords, referred to herein as their signature (page 3, paragraph 0030).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made, having the teachings of Tokuda, Glover and Agrawal before him/her, to modify Tokuda edition means for adding or deleting an arbitrary term with respect to the term list because that would provide a means for designing vastly enhanced searching, browsing and filtering systems as taught by Agrawal (page 1, paragraph 0009).

As to claims 8 and 17

Tokuda does not explicitly teach the claimed limitation "characterized in that said large classification generation module terminates repetition of the clustering processing when no cluster is obtained from the transformed DT matrix in the clustering processing".

Agrawal teaches each of the other second level topics may be divided at the third level to further topics. Also, in a similar fashion, further levels under the third level may be included in the topic hierarchy, or taxonomy. The final level of each path in the taxonomy terminates at a terminal or leaf node (page 6, paragraph 0087). Large sub-trees in the topic tree can be eliminated forthwith if the score of the root of those sub-trees are very poor (page 8, paragraph 0131).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made, having the teachings of Tokuda, Glover and Agrawal before him/her, to modify Tokuda terminates repetition of the clustering processing because that would provide a means for designing vastly enhanced searching, browsing and filtering systems as taught by Agrawal (page 1, paragraph 0009).

As to claims 9 and 18, Tokuda teaches the claimed limitations:

“Characterized by further comprising large classification label generation module for, if a virtual representative document is contained in a given cluster of clusters obtained by the clustering processing” as a new efficient supervised document classification procedure, whereby learning from a given number of labeled documents preclassified into a finite number of appropriate clusters in the database, the classifier developed will select and classify any of new documents introduced into an appropriate cluster within the classification stage (column 2, lines 22-28).

Tokuda does not explicitly teach the claimed limitation “generating a label of the cluster on which the virtual representative document is based from a term strongly connected to the virtual representative document”.

Glover teaches the virtual document classifier comprises the learning algorithm (not shown) that accepts as input a set of labeled input virtual documents. From the labeled input virtual documents the learning algorithm generates a prediction rule. After the virtual document classifier 106 is trained, a new unlabeled virtual document can be evaluated by the prediction rule to predict its label (page 4, paragraph 0031).

Also, Agrawal teaches that with reference to the hierarchy represented, statistics are calculated for the science node, based on the terms in all of the documents from the collection set that are classified in classes represented by nodes below the science node. Including the nodes labeled biology, chemistry, electronics, and all children nodes of those nodes (page 6, paragraph 0093).

Large sub-trees in the topic tree can be eliminated forthwith if the score of the root of those sub-trees are very poor. Text database population is not the only application of fast multi-level classification. With increasing connectivity, it will be inevitable that some searches will go out to remote text servers and retrieve results that must then be classified in real time (page 8, paragraph 0131).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made, having the teachings of Tokuda, Glover and Agrawal before him/her, to modify Tokuda strongly connected to the virtual representative document because that would provide a system which is sufficiently fast as taught by Agrawal (page 2, paragraph 0025).

Conclusion

4. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office Action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Contact Information

Any inquiry concerning this communication or earlier communications from the examiner should be directed to James Hwa whose telephone number is 571-270-1285. The examiner can normally be reached on 8:00 – 5:00. If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Don Wong can be reached on 571-272-1834. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only, for more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the PAIR system contact the Electronic Business

Application/Control Number:
10/563,311
Art Unit: 2163

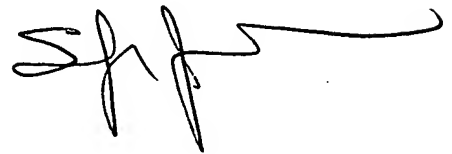
Page 14

Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

JH
01/25/2008

cy

James Hwa
Examiner
Art Unit 2163

A handwritten signature in black ink, appearing to be 'JHwa', with a long horizontal flourish extending to the right.A handwritten signature in black ink, appearing to be 'Wilson Lee', written in a cursive style.

WILSON LEE
PRIMARY EXAMINER

DETAILED ACTION

1. Applicant has amended claims 1, 3-6, 8-10, 12-15, 17 and 18 in the amendment filed on 11/21/2007. Claims 2, 7, 11 and 16 are canceled. Claims 1, 3-6, 8-10, 12-15, 17 and 18 are pending in this Office Action.

Response to Arguments

2. Applicant's arguments in regards to the rejections to claims 1 and 10 under 35 U.S.C. 103(a), have been fully considered. Applicant seems to argue the claims as amended. Consequently, new grounds of rejection are set forth bellow as necessitated by Applicant's amendment.

Amended claims now include "large classification generation module for generating a large classification of documents from each document in a bottom-up manner by repeatedly performing hierarchical clustering processing of setting a DT matrix generated by said DT matrix generation module in an initial state, causing said virtual representative document generation module to generate a virtual representative document for each cluster on a transformed DT matrix generated from the DT matrix by said DT matrix transformation module, generating a new DT matrix used for next hierarchical clustering processing by adding the virtual representative document to the transformed DT matrix and deleting documents belonging to the cluster of the virtual representative document from the transformed DT matrix, and outputting, for said each cluster, information associated with the documents constituting the cluster as large classification data" in claims 1 and 10. This added limitation distinguishes over the original claim by adding "repeatedly performing hierarchical clustering processing of

setting a DT matrix generated by said DT matrix generation module in an initial state".

As such, the additional also overcomes the rejection under 35 U.S.C. § 103 as being anticipated by Tokuda and Kauffman.

Despite the addition, the claims remain unpatentable for the reason now clearly set forth in the new ground of rejection below.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g) prior art under 35 U.S.C. 103(a).

3. Claims 1, 3-6, 8-10, 12-15 and 17-18 are rejected under 35 U.S.C. 103(a) as being unpatentable over Tokuda et al. (US Patent No. 7,024,400 B2, hereinafter "Tokuda") in view of Glover (US Patent Application No. 2003/0221163 A1, hereinafter "Glover") and Agrawal et al. (US Patent Application No. 2001/0037324 A1, hereinafter "Agrawal").